100,000 Genomes Project: opportunities and challenges

NHS-HE forum, 26th November 2014

Dr Clare Turnbull MA MSc PhD MRCP

Clinical Lead for Genomics England Cancer Programme

Senior Lecturer in Genomic Medicine, QMUL

About the 100K Genome Project

We are a new company set up by the Department of Health to help deliver the 100k Genome Project first announced by the Prime Minister David Cameron in December 2012.

This project will sequence the personal DNA code – known as a genome – of up to 100,000 patients over the next five years. This unrivalled knowledge will help doctors' understanding, leading to better and earlier diagnosis and personalised care. Based on expert scientific advice, we will start by tackling cancer, rare diseases and infectious diseases.

The company will manage contracts for sequencing, data linkage and analysis, and set standards for patient consent.

Prime Minister, David Cameron

"The UK will become the first ever country to introduce this technology in its mainstream health system."

Genomics England was announced by Jeremy Hunt, Secretary of State for Health, as part of the NHS 65th birthday celebrations on 5 July 2013.

He said: "The NHS has a long track record as a leader in medical science advances and it must continue to push the boundaries by unlocking the power of DNA data.

"The UK will become the first ever country to introduce this technology in its mainstream health system – leading the global race for better tests, better drugs and above all better, more personalised care to save lives.

"Genomics England will provide the investment and leadership needed to dramatically increase the use of this technology and drive down costs."

Secretary of State for Health, Rt Hon Jeremy Hunt MP

Genomics england

Genomics England – the mission

- **100,000 WGS** on NHS patients
  - Rare diseases
  - Cancer
  - [and pathogens]
- Working with NHS, academics and industry to make the UK a **world leader in Genomic Medicine**
- **Transformation** of NHS so that NGS can become routine investigation
- Generate **health and wealth**
- Leave a **legacy** of infrastructure, human capacity and capability

Genomics england

Rationale

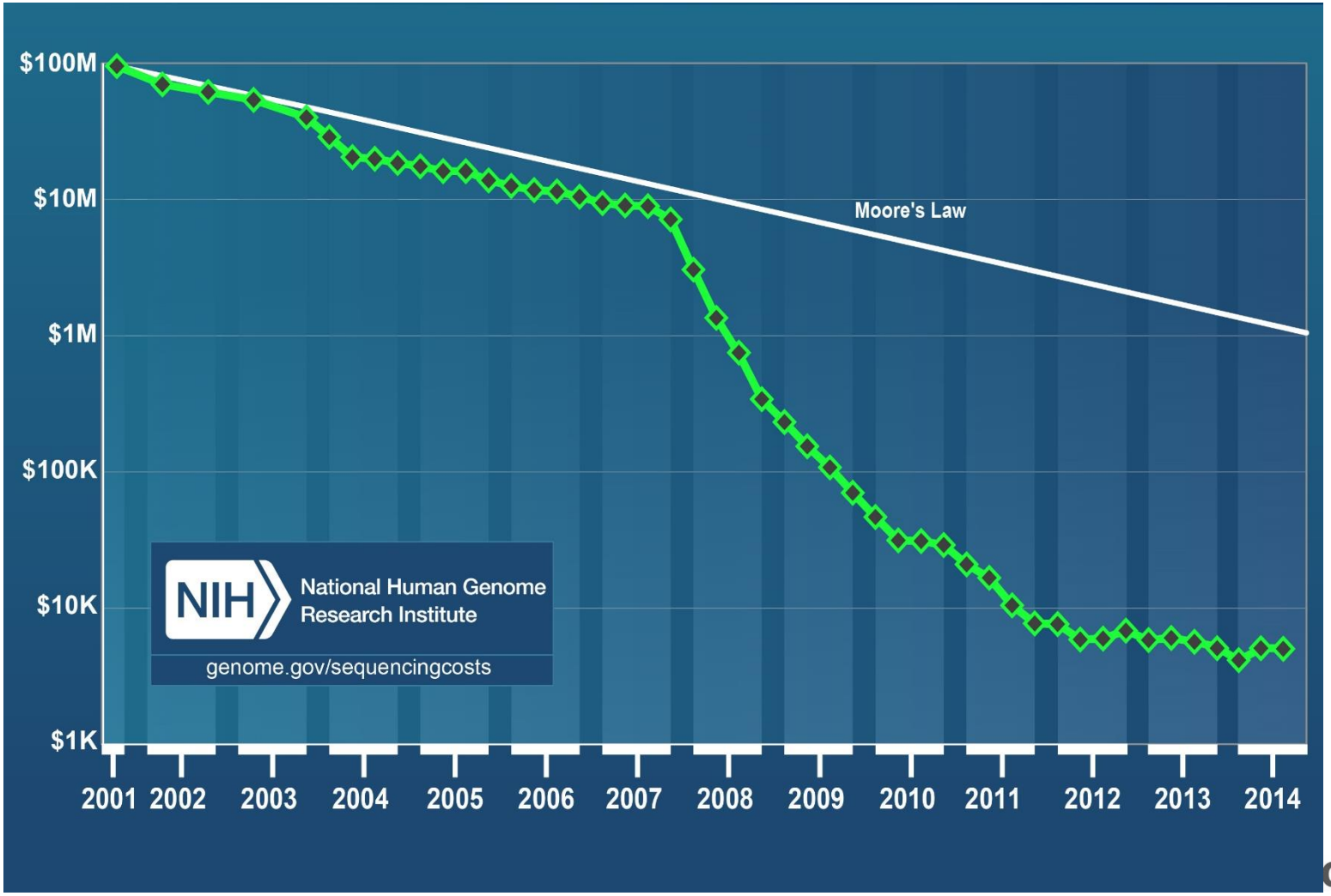
For Whole Genome Sequencing

Genomics england

Exomes vs Genomes

Data type	Large-scale structural changes	Balanced translocations	Distant consanguinity	Uniparental disomy	Novel / known coding variants	Novel /known non-coding variants
Targeted gene sequencing	No	No	No	No	Yes	No
SNP arrays	Yes	No	Yes	Yes	No	No
Array CGH	Yes	No	No	No	No	No
Exome	Partial	No	Partial	Partial	Yes	No
Whole genome	Yes	Yes	Yes	Yes	Yes	Yes

Cost of Genome sequencing



omics england

Rationale

For WGS for rare diseases

Genomics england

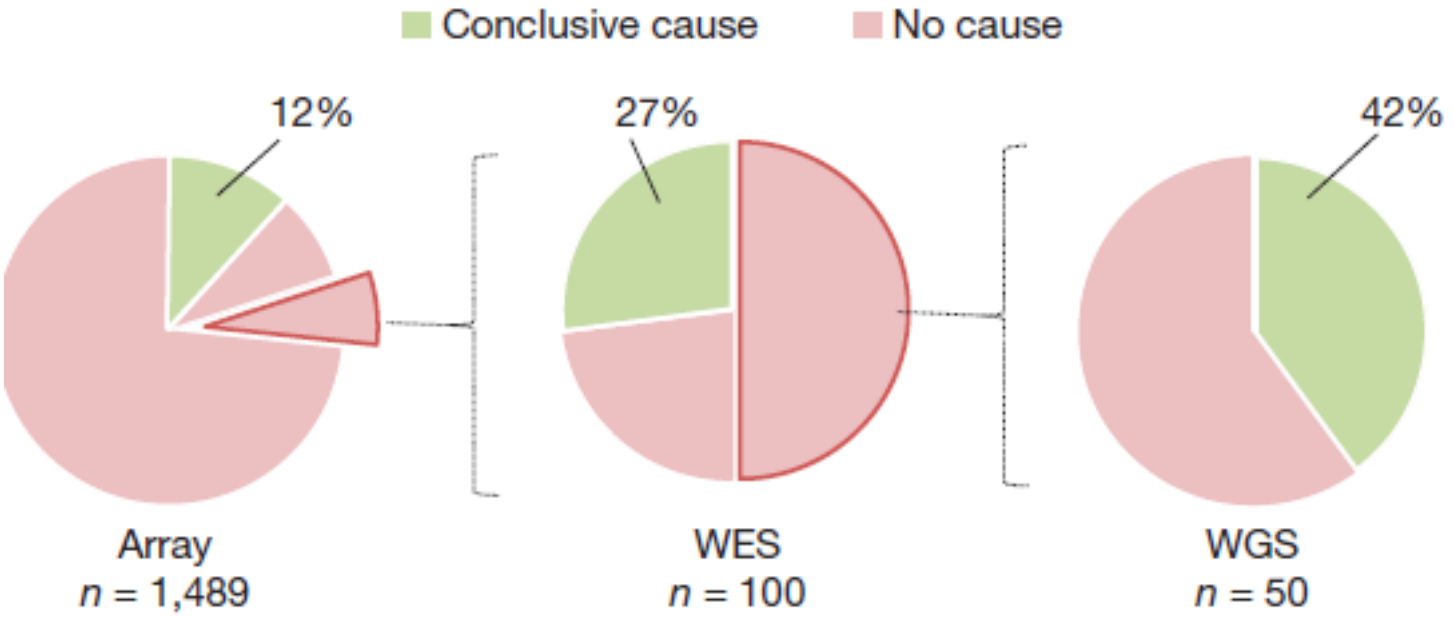
Why Rare diseases?

- Individual disease frequency of <1/2000 of population
- Over 7000 rare diseases
- Affect 2-5% of the UK population
- ~3 million people in the UK
- Diseases often disabling, shorten life, costly
  - ~75% affect children
  - ~30% die by their 5th birthday
- Early diagnosis and intervention may avoid disability; reproductive options for parents
- ~ 85% have a single gene defect

Genomics england

Rare Diseases - WGS

- Severe intellectual disability occurs in 0.5% of newborns
- Whole-genome sequencing at 80x in 50 parent-offspring with no diagnosis for their severe intellectual disability
- Overall 62% increase in diagnostic yield with WGS



Veltman et al Nature 2014

Genomics england

Genomics England – Rare Disease Partnerships

- NIHR Translational Research Collaborative in Rare Disease
  - £20m for deeper phenotyping
- NIHR BioResource in Rare Disease
  - 12,000 NHS patients with samples eligible for WGS
- NHS Clinical Genetics Service
- Engagement in large-scale consortia
  - UK Pfizer Rare diseases consortium
  - International Rare Diseases Research Consortium (IRDiRC)
  - European Union Committee of Experts on Rare Diseases (EUCERD)
- Decipher & Deciphering Developmental Disorders

Genomics england

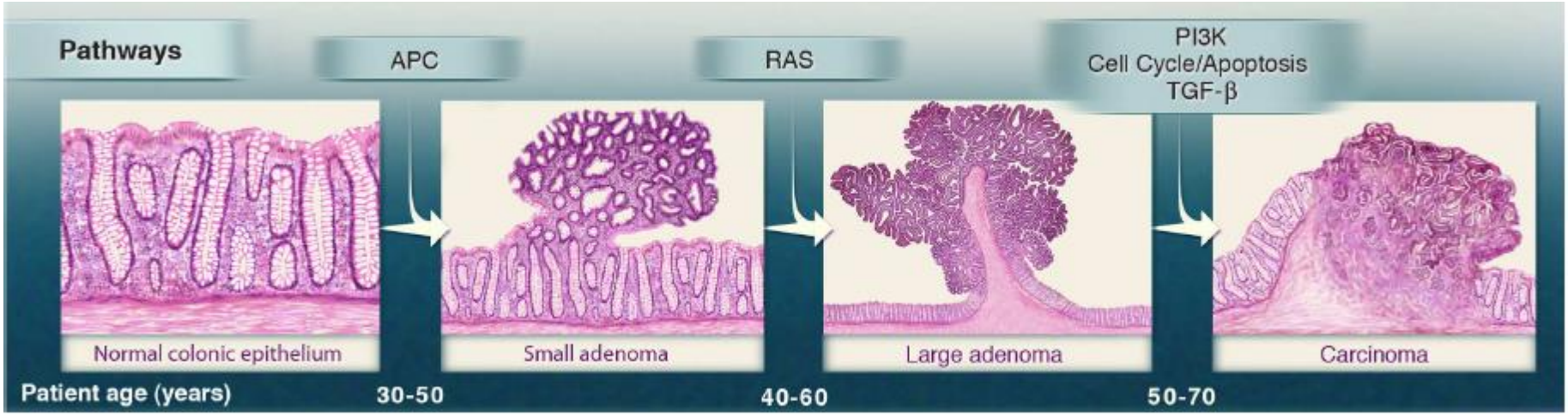
Rationale

For WGS for cancer

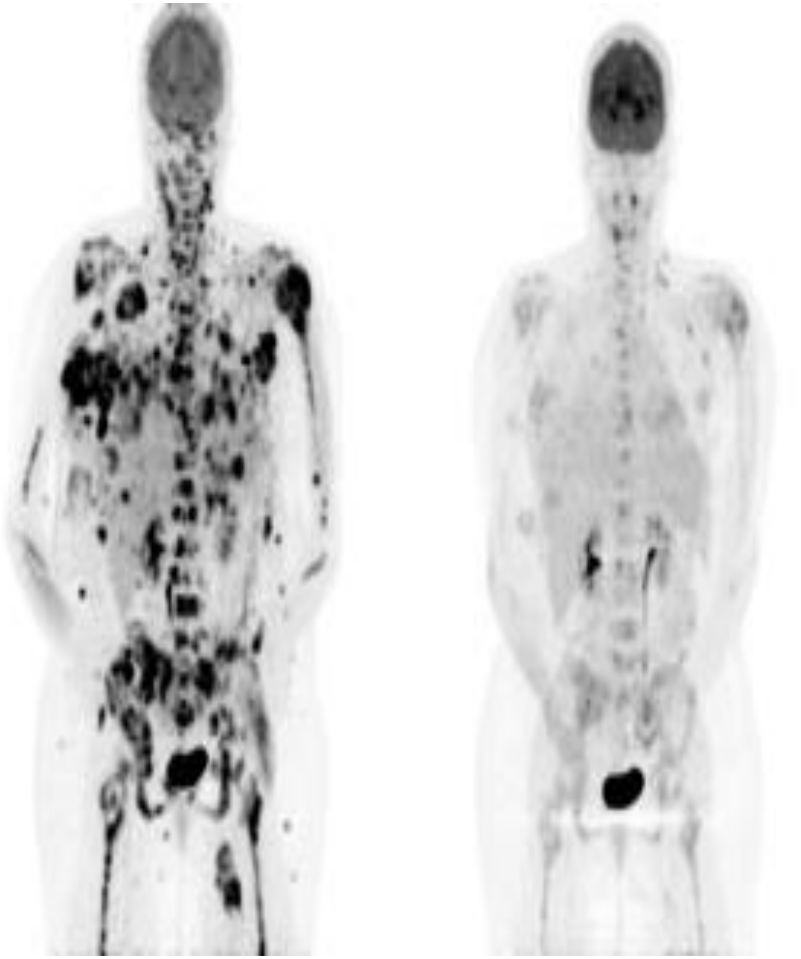
Genomics england

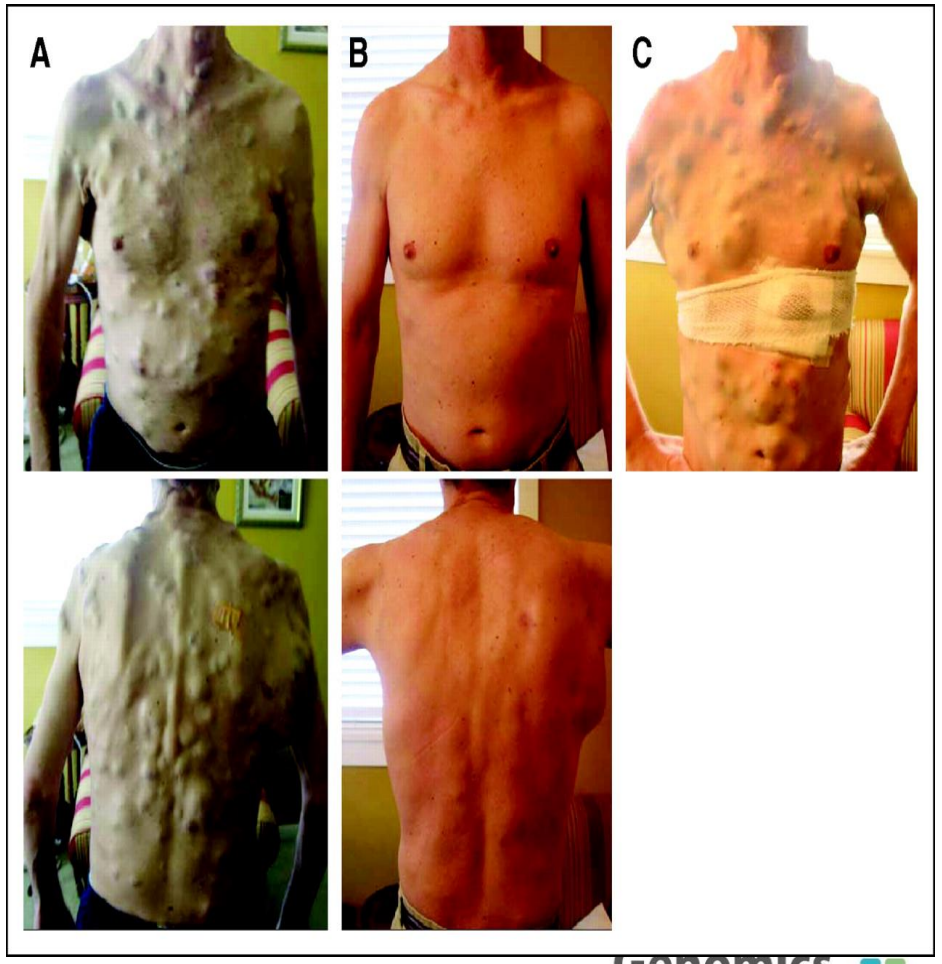
Cancer

- Cancer is a disease of disordered genomes
  - over 200 drivers known
  - drugs target mutations
  - Stratified medicine
  - Tumour heterogeneity/ evolution of cancer



Vemurafenib for Melanoma

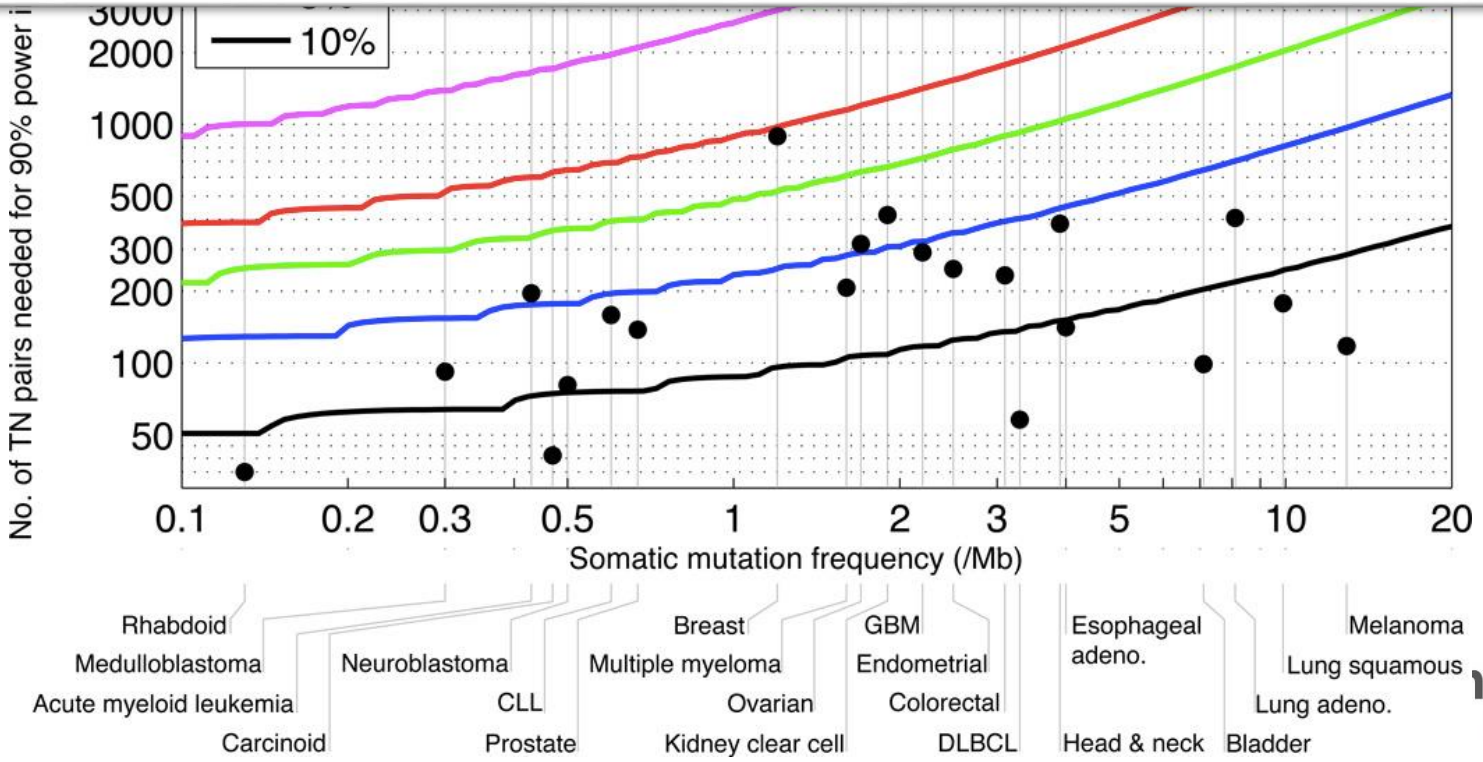




Genomics england

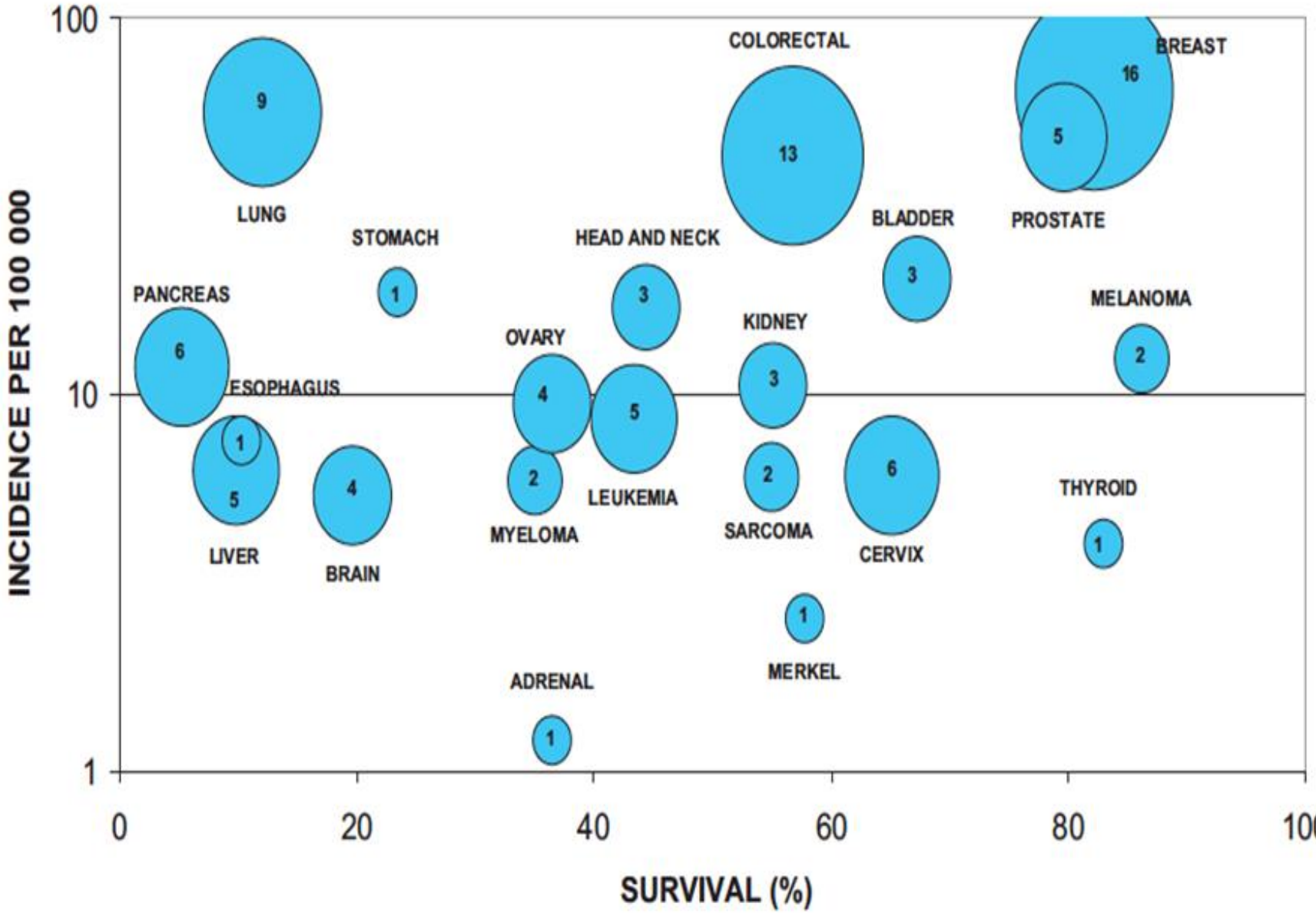
The case for lots of genomes

Creating a reasonably comprehensive catalogue of candidate cancer genes mutated in ≥2% of patients will require between approximately 650 samples (for tumours with ~0.5 mutations per Mb, such as neuroblastoma) to approximately 5,300 samples (for melanoma, with 12.9 mutations per Mb).



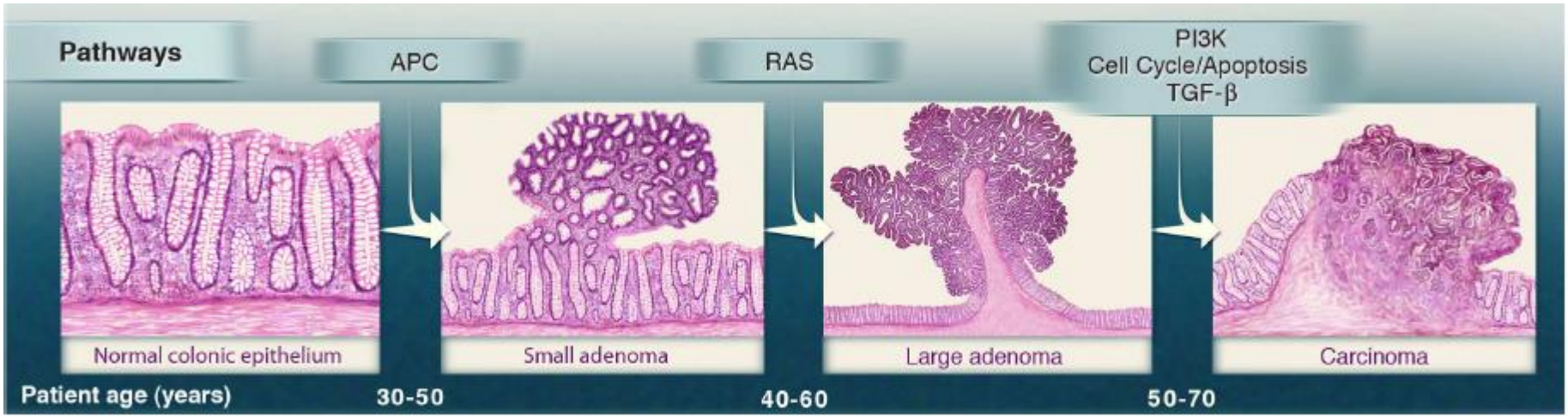
Lawrence et al, Nature 2014

Incidence v survival at 5 years



Cancer

- Cancer is a disease of disordered genomes
  - over 200 drivers known
  - drugs target mutations
  - Stratified medicine
  - Tumour heterogeneity/ evolution of cancer
- Cancers currently included in 100KGP:
  - Lung, breast, colon, prostate, ovary, CLL
- Collaborations:
  - CRUK Stratified Medicine/NIHR BRCs/Leukaemia consortia
  - MRC/NIHR Phenome Centres
  - Experimental Cancer Medicine Centres
  - International Cancer Genomes Consortium/The Cancer Genome Atlas



Genomics England and 100,000 Genomes programme

Establishing the programme

Genomics england

Key Partnerships in Establishment Phase

- Partnership of QMUL, UCL, The Farr, Oxford Big Data, UK Biobank, Sanger, EBI, Cambridge, Kings
- **Illumina** Partnership
- **Wellcome Trust:** award to establish NHS Genome Sequencing Centre (£27m)
- **MRC:** award to establish the infrastructure for UK Genomic Medicine Data Architecture (£24m)

PARTNERSHIP & INNOVATION WITHOUT REINVENTION

Genomics england

Genomics Medicine Centres

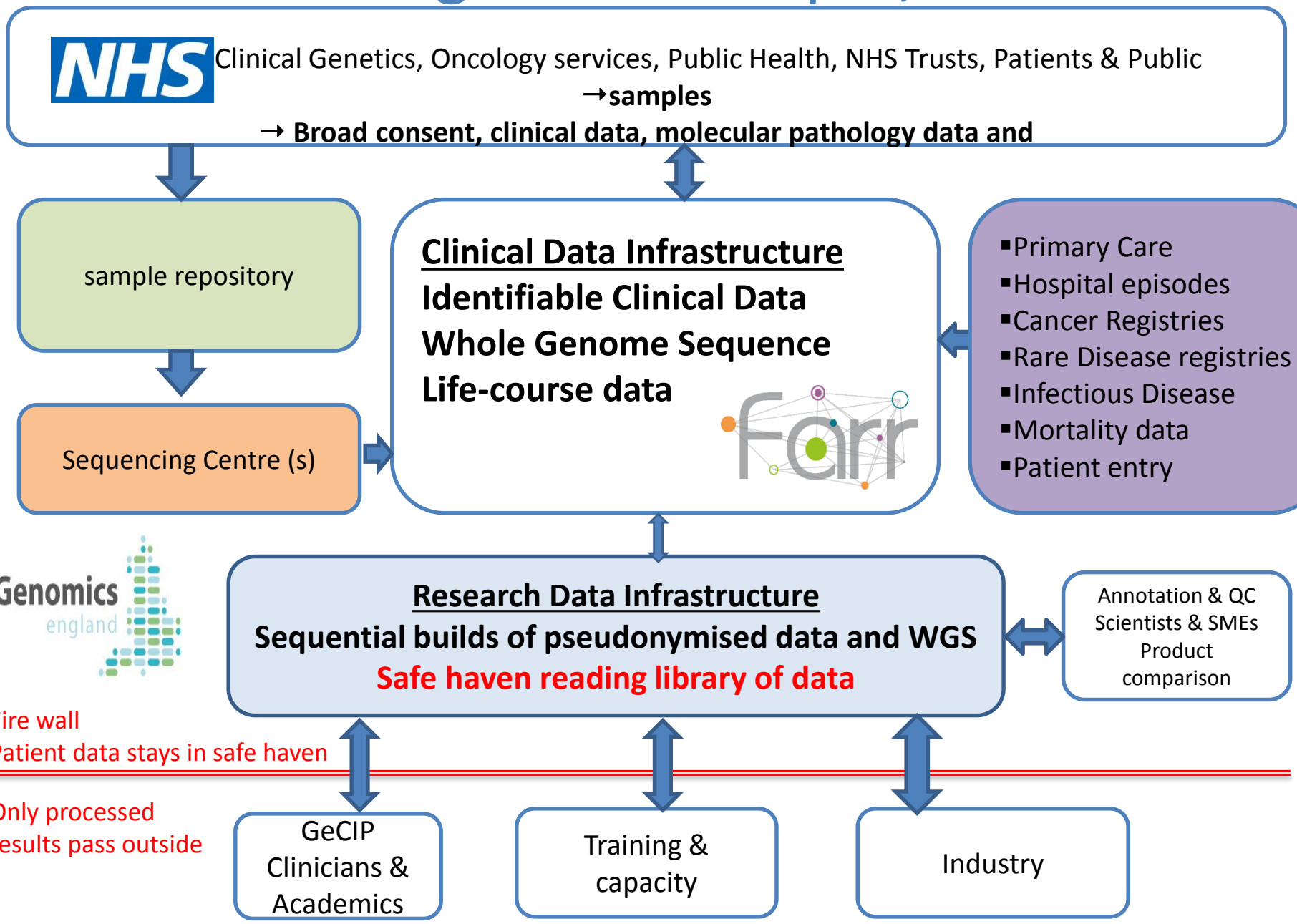
INVITATION TO TENDER

STAGE ONE: PRE-QUALIFICATION

NHS GENOMIC MEDICINE CENTRE SELECTION

- Partnership with NHS organisations
- Local NHS Lead and extended team
- ITT for contracts: consent, clinical data and quality DNA
- Can access to data on their samples
- Genomics England owns the data

Genomics England – sample, data flow



Ethics

- **Ethics Advisory Group:** Professor Mike Parker
- **For pilots:** Adapted NIHR BioResource and CRUK consent used
- **For main programme:**
  - NHS Genomic Medicine Protocol and Consent established through wide consultation
  - Protocol and consent processes under final agreements with REC
    - Return of secondary (looked-for) findings for high penetrance mutations and bi-parental findings with reproductive implications
    - Release of data to third parties
    - Prospective data linkage

Genomics england

Education in Genomic Medicine

NHS
Health Education England

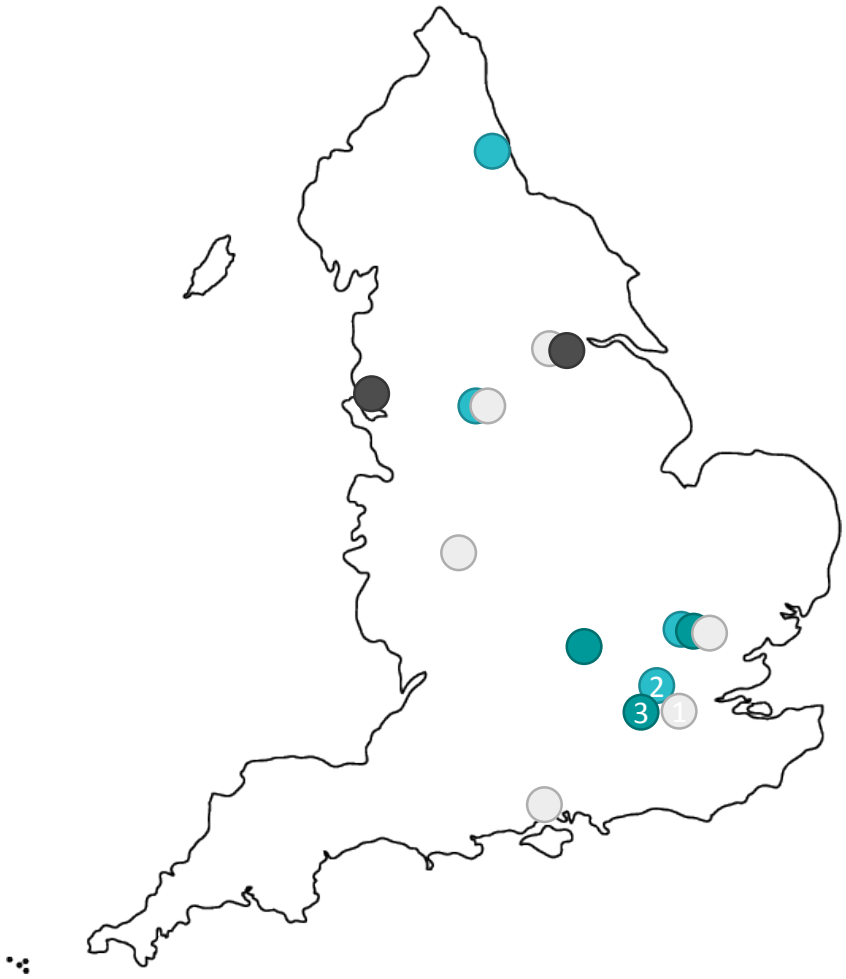
- **Workforce development in Genomic Medicine**:
  - specialist genetics, pathology and specialist clinical workforce
- **Specialized scientific training:** fellowships funded over 3/5 years:
  - Molecular Pathology including Infections and Pathogens
  - Genetics / genomics
  - Bioinformatics
- **MSc in Genomic Medicine**
  - CPD access to MSc modules for specialist practitioners
- S**pecialist on-line learning**
- B**ioinformatics workshops**

100,000 Genomes Project

Pilots and getting started

Genomics england

Genomics England Pilot Sites



Rare disease pilot
- Guys & St Thomas
- Manchester
- UCL Partners (Great Ormond Street, Moorfields)
- Newcastle
- Cambridge

Cancer (BRC) pilot
- UCL
- Imperial
- Guys & St Thomas
- Oxford
- Cambridge
- Southampton

Cancer (CRUK) pilot
- Leeds
- Manchester
- Birmingham
- Royal Marsden
- Cambridge
- Southampton

Cancer (CLL) pilot
- Leeds
- Liverpool

Genomics england

Genomics England Clinical Interpretation Partnership (GeCIP)

How clinicians and academics get more involved

Genomics england

GEL, GMCs and GeCIP

GEL

- Private Company (DoH)
- Sequencing capacity and infrastructure
- Informatics infrastructure
- Development of skills
- Delivery of 100,000 genomes
- Own IP

Genomics Medicine Centres

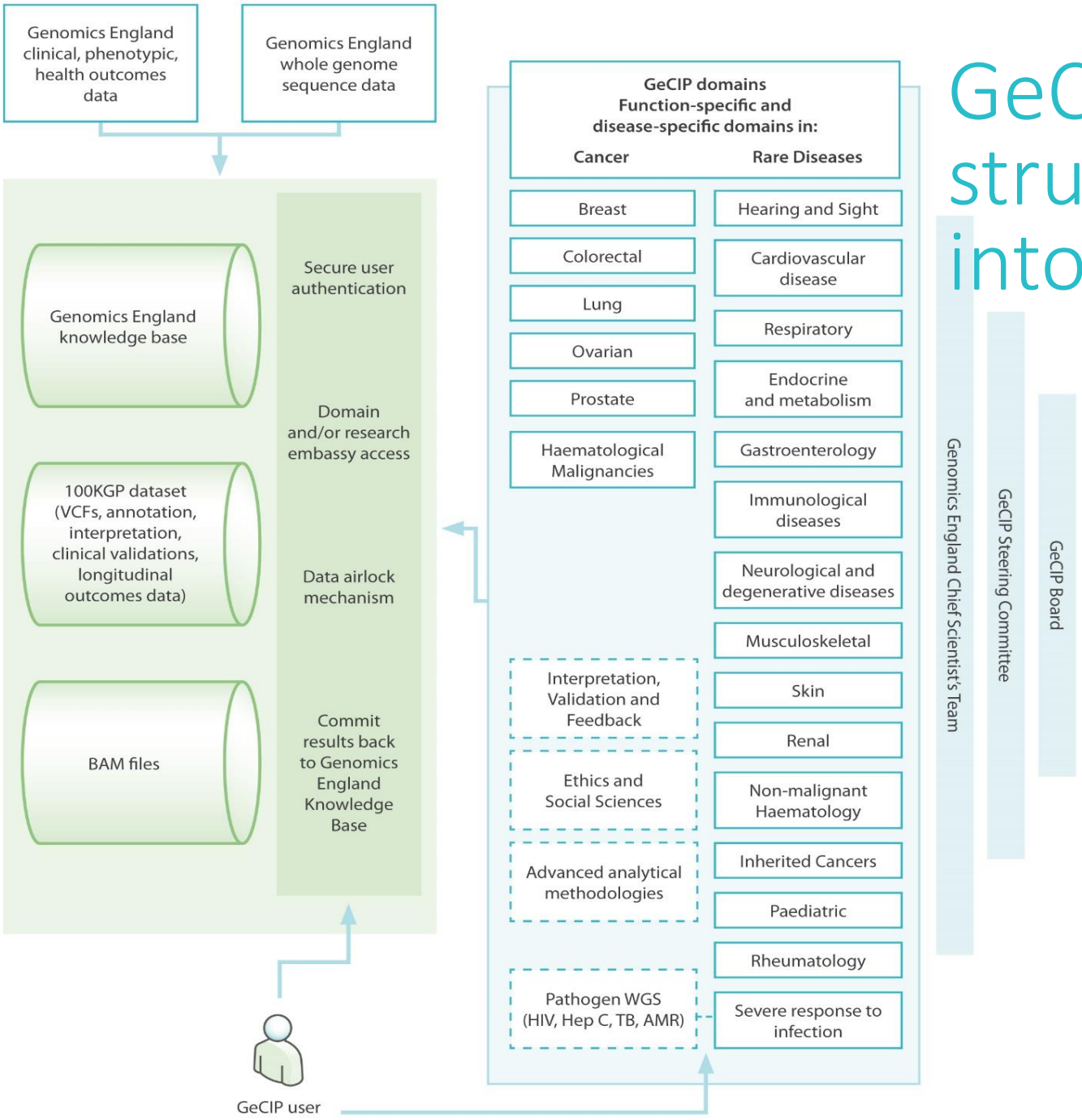
- NHS organisations
- Recruitment of patients
- Delivery of high quality samples
- Delivery of high quality phenotypic data
- Clinical validation of results
- Return of results to patients

GECIP

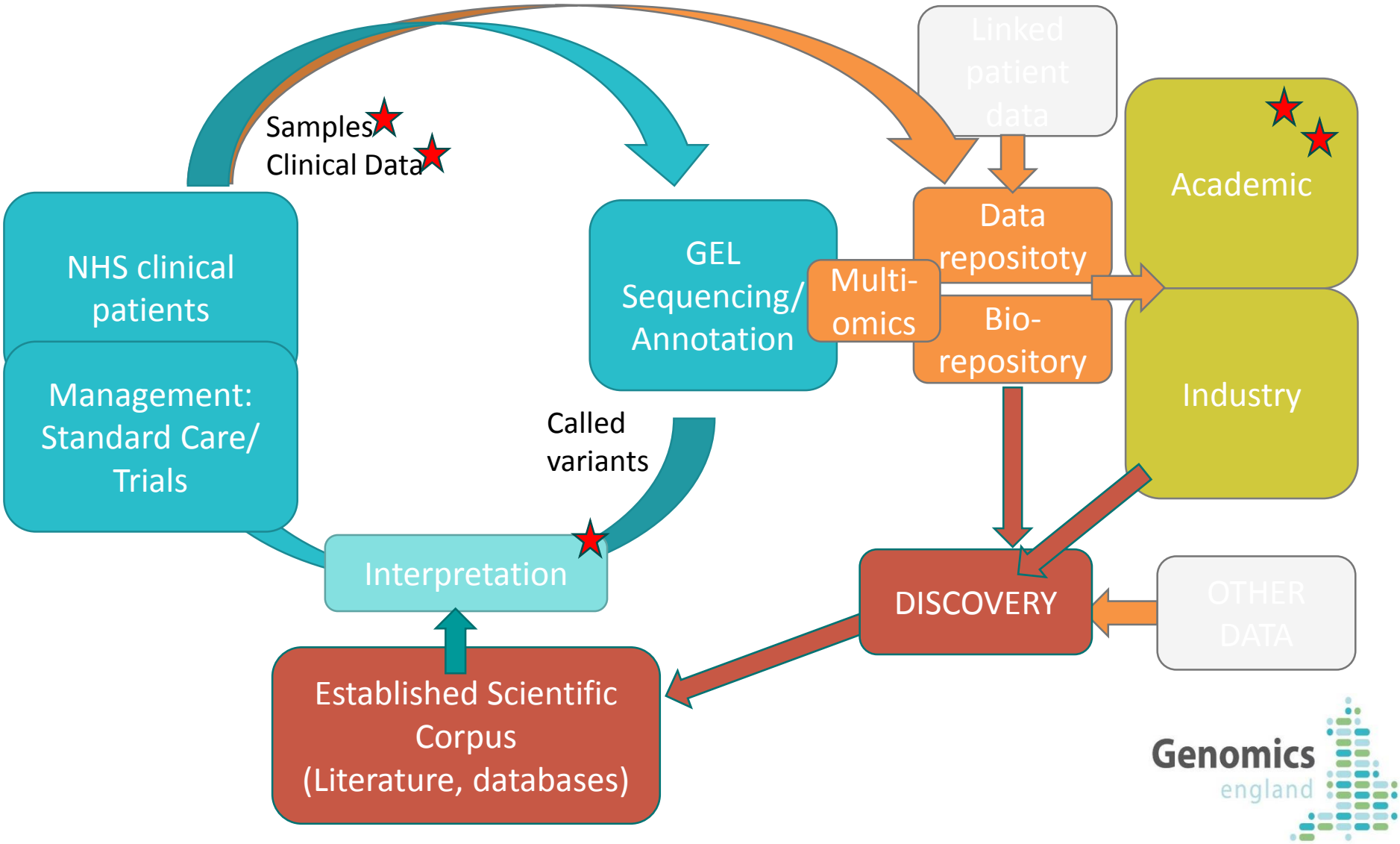
- Collaborations of clinicians and academics
- Enhanced clinical interpretation
- Research
- Research Training

Genomics england

GeCIP: structured into domains



GeCIP: where the experts add value



Genomics england

The challenges and opportunities

Genomics england

Data challenges 1: Genomic data

- **Genomic data**
  - Huge datasets. Compression
  - Multiple types of data (FASTQ, BAM, gVCF, VCF) ?store
  - Metadata on many levels
- **Analysis of Genomic data**
  - Evolution of analytical tools. Evaluation and validation
- **Data protection**
  - Pseudo-anonymisation
  - 3rd-party access
  - Linkage of familial genomic data
    - Key for clinical and research applications

Genomics england

Data challenges 2: Interpretation

- Our genomes are highly variant
- Each ‘healthy person’ has (in our exome)
  - 22,000 genetic variants
  - 9,000 non-synonymous genetic variants
  - 100 rare non-synonymous genetic variants
  - 5 rare protein-truncating genetic variants
  - 0-2 de novo mutations
- The ‘rules’ of interpretation vary gene-to gene

  →Interpretation of exome data is complex, iterative and expert (know your genes, know your diseases)
- The exome is about 1.5% of the genome….

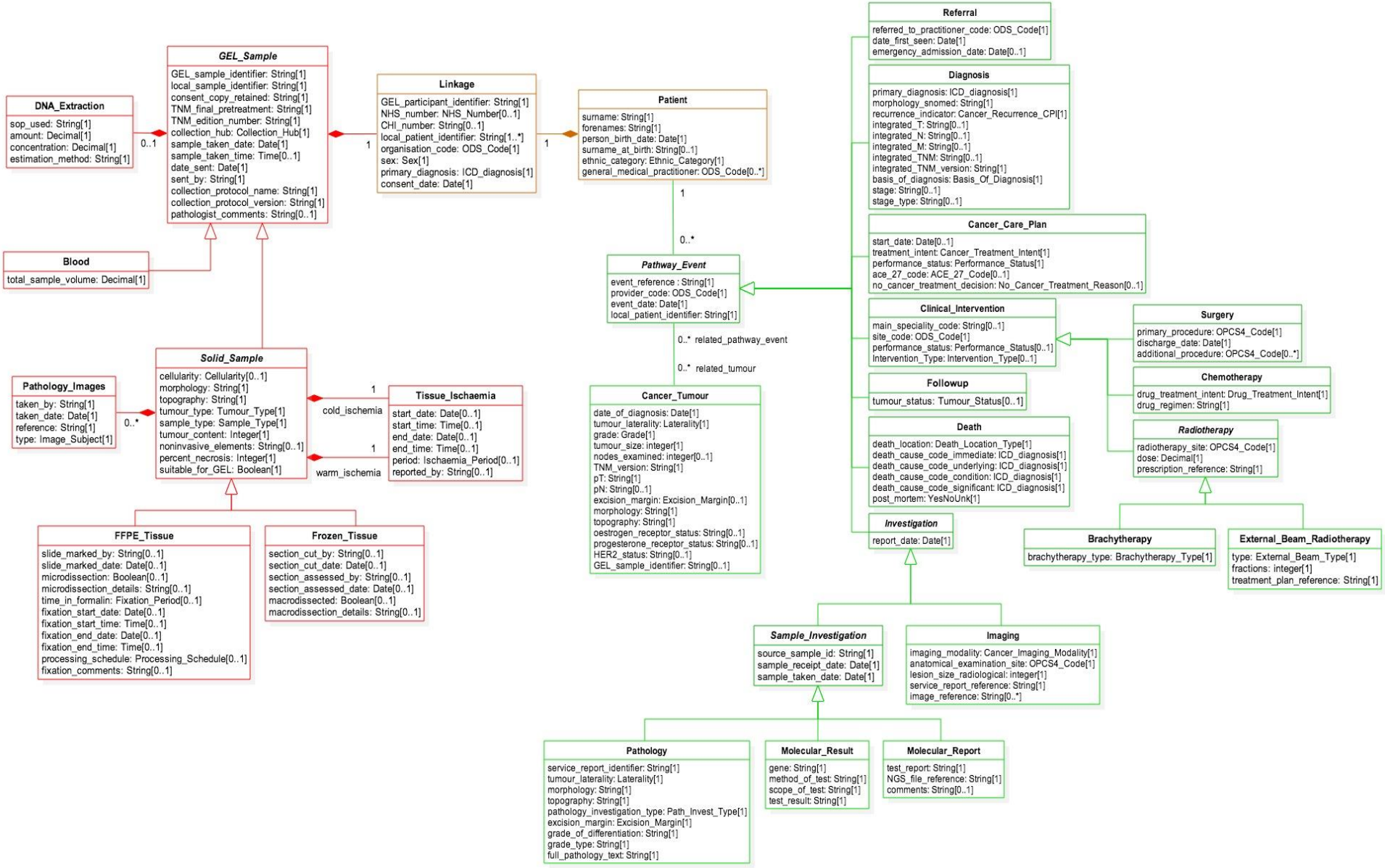
Genomics england

Data challenges 3: Clinical data capture models

- **Data sources**
  - Prospective data (patients coming through NHS)
  - Retrospective data (clinical trials)
- **Hierarchies within models**
  - Patient
  - Cancer
  - Sample for WGS
  - Clinical episode
- **Existing models for clinical data capture**
  - NCIN-NCRS: COS-D, SAC-T, RTDS
- **Rare diseases**
  - Over 150 data individualised capture tools required (one per disease)
  - Massively complex disease ontologies: commonality and plain English
- **Flexibility vs universality of structure**
- **Identifiers for data linkage** (NHS number, ?linking families)
- **NHS==multiple different localised IT systems**
  - Over-worked
  - Under-resourced

Genomics england

Data Model:
Generic Model for solid tumours

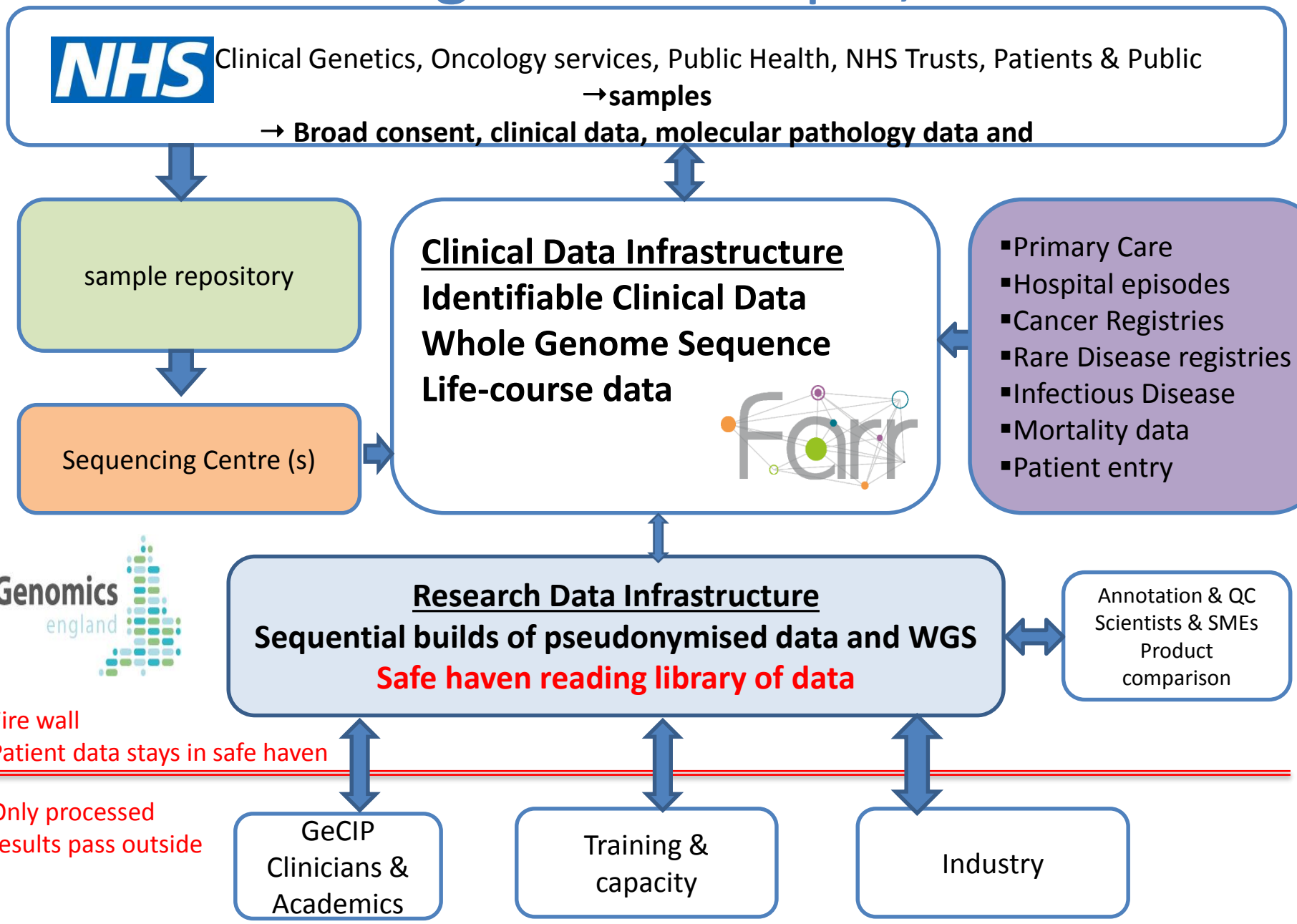


Data challenges 4: Data access and capture

- Safe haven storage of 100KGP dataset
  - '**Embassies'** for data GeCIP domains for data viewing and analysis
- Levels of access
  - Which parts of 100KGP dataset
  - Level of access: anonymised
- Accelerated access for GeCIP domains: quid pro quo
- Matrix approach offers challenges
- Release of data: summary level only
- Accrual of discovery in **Genomics England Knowledge Base**

Genomics england

Genomics England – sample, data flow



Other Broad Challenges

- Ethical conservatism and genetic exceptionalism
- Engagement of NHS staff
- Engagement of patients
- Standardisation of sample collection
  - Surgical processes
  - Pathology pipelines
- Implementation of this complex technology in real-time patient care (2 weeks for cancer)

Genomics england

Opportunities

- Efficiency of genomic testing strategy
  - Multiple samples, multiple labs, iterative journey of incomplete genomic information
- Outcomes of genomic testing
  - Diagnoses
  - Stratified medicine
- NHS as a realtime research cohort
  - Data linkage
  - Generate rich lifelong cohort data
  - Rich sample repository
- Streamlining of research activity

Genomics england

Summary

- **100,000 WGS** on NHS patients and pathogens
- Working with NHS, academics and industry to **drive Genomic Medicine into the NHS**
- Support that with **education**
- **Transformation** of NHS: consenting, pathology pipelines, clinical data collection, applying genomic data in the clinic
- Leave a **legacy** of NGS Centres, sample pipeline and biorepository, large-scale data store that makes this usable by the NHS
- New diagnostics and therapies and opportunities for patients
- Aware of the challenges
- All by the end of 2017…..

Genomics england

Put genomic medicine right into the National Health Service; have the National Health Service as the central hub for genomic research

Genomics England – who are we?

Officers

- Sir John Chisholm (Executive Chair)
- Mark Caulfield (Chief Scientist)
- Nick Maltby (Company Secretary)
- Jim Davies (Informatics)
- Viv Parry (Outreach)

Board

- Prof Dame Sally Davies (CMO)
- Kevin Dean (Cisco)
- Prof Sir John Bell
- Jon Symonds (Audit)
- Prof Sir Malcom Grant (NHSE)

Advisory Committees

- Science: Sir John Bell,
- Informatics: Kevin Dean
- Ethics: Mike Parker

Genomics england


Genomics
england